

## **Scale Score Interpretations and Limitations**

The scale scores associated with the SBAs are not vertically equated. Therefore interpretation of individual score differences between the assessments is not appropriate.

Because the scale score are established independently by grade/subject combination, a comparison of scale scores between subjects is also inappropriate. Each scale score is based on a set of standards that define that content area and on items that operationalize that definition. The appropriate comparison is to compare the student to summary statistics for other students.

## **STANDARDS VALIDATION**

Traditionally, standard setting methods have fallen into two camps: test-centered methods and examinee-centered methods (Jaeger, 1989). With test-centered methods, the standard-setting judgment is made primarily by referencing the test itself, most often based upon an inspection of the actual test items. Examinee-centered methods call for judgments to be made about the performance of examinees, such as the performance of established mastery and non-mastery groups. As Kane (1995) points out, “all standard setting is based on judgments.” Performance standard setting uses a process from which the best judgments are obtained from the people in the best position to make those judgments. Typically, these are content experts, people familiar with the skills and knowledge to be learned.

### **Standard Setting Panels—Alaska Teachers and Stakeholders**

Representatives of three broad groups typically comprise the standard setting panel. These groups include teachers, non-teacher educators, and non-educators. Non-educators are drawn from civic and business leaders. The non-teacher educators could be curriculum directors or administrators. Its selection process also considers size and location of districts and schools, socioeconomic conditions, and other demographics.

While there is no simple answer to the question of how many panelists are required to establish reliable standards, the larger the number involved the more confidence there can be that the result will generalize to another selection of panelists. The consensus in the literature indicates samples of 20 or more are adequate for establishing stable performance standards. The current plan sought 24 panelists for each content area or a total of 72. Separate panels were used for mathematics, reading, and writing.

## Modified Bookmark Procedure

Because there was a need to maintain consistency during the transition from the old Benchmark assessments to the new assessment system a Modified Bookmark Procedure (Lewis, Mitzel, & Green, 1996) was utilized to set the cut scores. The primary modification was that participants were provided with information as to where the bookmark would be placed if EED were to apply equipercentiles from the 2004, grade 6 Benchmark administration prior to beginning Round One. Participants were then instructed to move the bookmark only if the GLEs warranted such a modification. A schedule for the Standards Validation is provided in Appendix 21.

## Recommended Cut Scores

Appendix 22 shows the Round by Round summaries for each grade and content area. Once the panelists' recommendation was calculated, the impact data was smoothed across grades within each subject area using a triweight kernel function (SPSS Inc, 1999):

$$f(x) = a(1 - (\frac{x}{h})^2)^3 \quad : (-h \leq x \leq h), \text{ else } 0,$$

where the constant  $a$  scales the formula as a probability kernel function and the constant  $h$  is the bandwidth.

The smoothed results are presented in Tables 8–1 through 8–3.

**Table 8–1. Percent of Students in Proficiency Categories – Mathematics**

<b>Mathematics</b>							
	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>Far Below Proficient</b>	12.5%	16.5%	15.1%	17.3%	15.9%	18.9%	23.2%
<b>Below Proficient</b>	12.0%	14.7%	18.1%	17.8%	23.1%	18.9%	20.3%
<b>Proficient</b>	44.3%	41.8%	34.9%	36.0%	37.6%	39.0%	32.0%
<b>Advanced</b>	31.2%	27.0%	31.9%	28.9%	23.4%	23.2%	24.5%

**Table 8–2. Percent of Students in Proficiency Categories – Reading**

<b>Reading</b>							
	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>Far Below Proficient</b>	9.2%	8.9%	8.1%	7.3%	8.2%	4.6%	3.8%
<b>Below Proficient</b>	11.7%	13.0%	14.4%	15.2%	16.4%	15.1%	18.8%
<b>Proficient</b>	44.6%	53.3%	52.5%	49.3%	48.0%	49.0%	40.4%
<b>Advanced</b>	34.5%	24.8%	25.0%	28.2%	27.4%	31.3%	37.0%

**Table 8–3. Percent of Students in Proficiency Categories – Writing**

<b>Writing</b>							
	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
<b>Far Below Proficient</b>	2.2%	1.4%	0.9%	3.9%	8.3%	6.8%	8.5%
<b>Below Proficient</b>	23.1%	22.3%	23.7%	24.6%	21.0%	20.0%	18.5%
<b>Proficient</b>	47.5%	54.7%	52.9%	45.0%	55.5%	66.2%	69.1%
<b>Advanced</b>	27.2%	21.6%	22.5%	26.5%	15.2%	7.0%	3.9%

## TRANSFORMATIONS

The student ability measures were transformed mathematically to a more convenient metric. To maintain consistency from administration to administration, the minimum scale scores necessary for each proficiency level are provided in Tables 8–4 through 8–6. Tables 8–7 through 8–9 provide the equations used for each transformation. These equations were applied to the overall test as well as to each reporting subscale.

**Table 8–4. Minimum Mathematics Scale Scores for Each Proficiency Level**

<b>Mathematics</b>									
<b>Grade</b>	<b>Raw Score Cut Point</b>			<b>Below Proficient</b>		<b>Proficient</b>		<b>Advanced</b>	
	<b>Below Proficient</b>	<b>Proficient</b>	<b>Advanced</b>	<b>SS Cut</b>	<b>SSSE</b>	<b>SS Cut</b>	<b>SSSE</b>	<b>SS Cut</b>	<b>SSSE</b>
3	25	33	51	263	17	300	17	390	20
4	27	35	51	260	18	300	18	383	22
5	25	35	50	252	17	300	17	373	20
6	25	34	49	258	17	300	17	376	19
7	22	33	49	248	18	300	17	383	20
8	25	34	50	258	18	300	17	379	20
9	26	36	50	258	17	300	17	370	20